

# Throughput and Efficiency of a Mass Spectrometry-Based Screening Assay for Protein–Ligand Binding Detection

Erin D. Hopper,<sup>a</sup> Petra L. Roulhac,<sup>a</sup> Michael J. Campa,<sup>b</sup>  
Edward F. Patz Jr.,<sup>b</sup> and Michael C. Fitzgerald<sup>a</sup>

<sup>a</sup> Department of Chemistry, Duke University, Durham, North Carolina, USA

<sup>b</sup> Department of Radiology, Duke University Medical Center, Durham, North Carolina, USA

An H/D exchange- and MALDI mass spectrometry-based screening assay was applied to search for novel ligands that bind to cyclophilin A, a potential therapeutic and diagnostic target in lung cancer. The assay is based on stability of unpurified proteins from rates of H/D exchange (SUPREX), which exploits the H/D exchange properties of amide protons to measure the increase in a protein's thermodynamic stability upon ligand binding in solution. The current study evaluates the throughput and efficiency with which 880 potential ligands from the Prestwick Chemical Library (Illkirch, France) could be screened for binding to cyclophilin A. Screening was performed at a rate of 3 min/ligand using a conventional MALDI mass spectrometer. False positive and false negative rates, based on a set of control data, were as low as 0% and 9%, respectively. Based on the 880-member library screening, a false positive rate of 0% was observed when a two-tier selection strategy was implemented. Although novel ligands for cyclophilin A were not discovered, cyclosporin A, a known ligand to CypA and a blind control in the library, was identified as a hit. We also describe a new strategy to eliminate some of the complications related to back exchange that can arise in screening applications of SUPREX. (J Am Soc Mass Spectrom 2008, 19, 1303–1311) © 2008 American Society for Mass Spectrometry

The speed and sensitivity of modern mass spectrometers make them attractive tools for high throughput screening (HTS) assays, and several mass spectrometry-based HTS assays have been developed in recent years [1–3]. We recently reported on a mass spectrometry-based assay for protein–ligand binding detection that utilizes an abbreviated version of stability of unpurified proteins by rates of H/D exchange (SUPREX), which is an H/D exchange- and mass spectrometry-based technique capable of detecting and quantifying protein–ligand binding interactions [4–9]. In the abbreviated version of SUPREX (referred to hereafter as single-point SUPREX), binding events are detected by measuring the target protein's mass change after H/D exchange in a deuterated buffer containing a specific concentration of a chemical denaturant [10].

Several inherent advantages of SUPREX make it particularly well-suited for an HTS assay. Unlike most radiometric or fluorescence-based assays, single-point SUPREX can be performed on protein–ligand complexes directly in solution without immobilization or labeling of the target or library compounds. The technique

does not require time-consuming separations or filtrations and is relatively general (i.e., it can be readily transferred to most protein–ligand systems). Additional advantages include the ability to study multicomponent mixtures and to make measurements on picomole quantities of protein.

The single-point SUPREX protocol was initially developed in a proof-of-concept study using the S-protein and a small test library of five peptides with known binding affinities for the S-protein [10]. The current study represents the first application of single-point SUPREX in a screening project designed to identify novel protein ligands, and as such it provides the first measure of the throughput and efficiency of this technique. The target protein for this study, cyclophilin A (CypA), is a protein that is overexpressed in lung tumor cells [11] and appears to be necessary for normal tumor growth [12]. Thus, tight binding ligands for CypA could potentially be used as diagnostic imaging agents as well as lung cancer therapeutics.

Cyclosporin A (CsA), an immunosuppressive drug, is the most well-studied and tightest-binding CypA ligand identified to date [13–17]. Unfortunately, biodistribution studies have shown that CsA is ill-suited for use as an imaging agent as it has been shown in rats to be rapidly taken up by the liver and excreted in the GI tract [18]. The immunosuppressive activities of CsA [13]

Address reprint requests to Professor M. C. Fitzgerald, Department of Chemistry, Box 90346, Duke University, Durham, NC 27708-0346, USA. E-mail: [michael.c.fitzgerald@duke.edu](mailto:michael.c.fitzgerald@duke.edu)

\*Current address: Department of Neurobiology, Duke University Medical Center, Durham, North Carolina, USA

also make it unattractive as a lung cancer therapeutic. One goal of the current work was to identify novel CypA ligands that might be more amenable to diagnostic imaging and therapeutic applications for lung cancer.

In an attempt to identify novel molecular scaffolds that bind to CypA, the 880-member Prestwick Chemical Library (Illkirch, France) was screened. The Prestwick Chemical Library contains a variety of structurally diverse compounds with known safety and bioavailability in humans. Over 85% of the compounds in the library are off-patent drugs that are marketed in a wide range of therapeutic areas. CsA was also present in the Prestwick Chemical Library, where it served as a blind control. This blind control was the only ligand identified in our assay as a hit. The focus of this report is to describe the analytical capabilities (i.e., the throughput and efficiency) of single-point SUPREX.

## Experimental

### Materials

The CypA (human) used in this work was obtained by recombinant DNA methods that involved the following steps: (1) overexpressing the CypA protein as a glutathione S-transferase (GST) fusion protein in *E. Coli* (BL21-DE3), (2) purifying the fusion protein using a GST-binding resin, (3) removing the GST tag in an overnight incubation with 5 units of thrombin per mg of fusion protein, and (4) removing the GST and thrombin with GST-binding resin (Promega Corp., Madison, WI) and HiTrap Benzamidinium FF (Amersham Biosciences, Piscataway, NJ), respectively. The 880-compound Prestwick Chemical Library was obtained from the Small Molecule Synthesis and Screening Facility in the Center for Chemical Biology at Duke University. The library compounds were provided as 1  $\mu$ L aliquots of 1 mM solutions in dimethyl sulfoxide (DMSO). CsA was purchased from LKT Laboratories (St. Paul, MN). Myoglobin (horse skeletal muscle), trypsin inhibitor (soybean), deuterium oxide ( $D_2O$ ; 99.9% atom D), sodium deuterioxide (40% by weight in  $D_2O$ , 99.9% atom D), and sinapinic acid (SA) were from Sigma (St. Louis, MO). Phosphoric acid- $d_3$  was purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA), and guanidinium chloride (GdmCl) was purchased from EMD Chemicals, Inc. (Gibbstown, NJ). Trifluoroacetic acid (TFA) was from Halocarbon (River Edge, NJ), and acetonitrile (ACN) and DMSO were from Fisher (Fairlawn, NJ).

### Theoretical SUPREX Curves

Equation 1, which was derived in reference [4] for the analysis of SUPREX data, was used to generate the theoretical SUPREX curves in this work.

$$\Delta\text{Mass} = \Delta M_{\infty} + (\Delta M_0 - \Delta M_{\infty}) e^{\frac{\langle k_{\text{int}} \rangle t}{(1 + K_{\text{fold}})}} \quad (1)$$

The variables in eq 1 are defined as follows:  $\Delta\text{mass}$  is the mass difference between the protonated and deuterated forms of the protein,  $\Delta M_0$  is the  $\Delta\text{mass}$  of the protein before the protein's globally protected amide protons are exchanged with solvent deuterons (i.e., the  $\Delta\text{mass}$  of the pre-transition baseline of a SUPREX curve),  $\Delta M_{\infty}$  is the  $\Delta\text{mass}$  of the protein after all of the protein's amide protons have exchanged with solvent deuterons (i.e., the  $\Delta\text{mass}$  of the post-transition baseline of a SUPREX curve),  $\langle k_{\text{int}} \rangle$  is the average intrinsic exchange rate for an unprotected amide proton,  $t$  is the H/D exchange time used in the SUPREX experiment, and  $K_{\text{fold}}$  is given by eq 2.

$$K_{\text{fold}} = e^{\frac{-(\Delta G_f + m[\text{GdmCl}])}{RT}} \quad (2)$$

The variables in eq 2 are defined as follows:  $R$  is the universal gas constant,  $T$  is the temperature,  $m$  is defined as  $\delta\Delta G_f / \delta[\text{denaturant}]$ ,  $[\text{GdmCl}]$  is the concentration of GdmCl in the exchange buffer, and  $\Delta G_f$  is the protein's folding free-energy in the absence of denaturant.

In using eqs 1 and 2 to construct the theoretical SUPREX curve for CypA in the absence of ligand,  $\Delta M_0$  and  $\Delta M_{\infty}$  were assigned values of 55 and 95 Da (respectively) based on the experimental data in reference [17], and  $\langle k_{\text{int}} \rangle$  was assigned a value of  $7.08 \text{ s}^{-1}$  based on a calculation using the program SPHERE, which performs the calculation using model dipeptide data along with the protein sequence and experimental temperature and pH values [19, 20]. We set  $t$  to 35 min, and  $m$  and  $\Delta G_f$  were assigned values of 3.7 kcal/(mol M) and  $-10.4 \text{ kcal/mol}$  (respectively) based on an average of the two sets of values reported for CypA in reference [17].

Construction of the theoretical curves for CypA in the presence of the hypothetical ligands required the calculation of a  $\Delta\Delta G_f$  value (i.e., the change in  $\Delta G_f$  upon ligand binding) according to eq 3 [21].

$$\Delta\Delta G_f = -nRT \ln \left( 1 + \frac{[L]}{K_d} \right) \quad (3)$$

In eq 3,  $[L]$  is the concentration of free ligand in the exchange buffer,  $n$  is the number of equivalent ligand binding sites in the protein,  $K_d$  is the dissociation constant of the hypothetical ligand, and  $R$  and  $T$  are defined above. In our calculations using eq 3,  $n$  was set to 1 and  $[L]$  was set to 90  $\mu\text{M}$ , which was the estimated ligand concentration in the SUPREX buffers. A  $\Delta G_f$  value was calculated for CypA complexed with each hypothetical ligand by adding the corresponding  $\Delta\Delta G_f$  value for CypA in the presence of each hypothetical ligand to  $-10.4 \text{ kcal/mol}$  (i.e., the  $\Delta G_f$  value of the unbound protein—see above). Ultimately the resulting  $\Delta G_f$  values were used in eq 1, as described above, to

generate theoretical SUPREX curves for CypA in the presence of each hypothetical ligand.

### *Buffer Preparation*

A deuterated exchange buffer solution (20 mM phosphate, pD 7.4) containing 1.5 M GdmCl was prepared. The pD of the buffer was adjusted by adding sodium deuterioxide while monitoring pH using a Jenco microcomputer pH-Vision 6072 pH meter equipped with a Futura calomel pH electrode from Beckman (Fullerton, CA). pH measurements were converted to pD measurements by adding 0.4 to the pH reading [22]. Before protonated GdmCl was used in the buffers, it was brought to a deuterium content of >99% by performing four cycles of dissolution in D<sub>2</sub>O (19 g GdmCl in 25 mL D<sub>2</sub>O) and lyophilization. The final concentration of GdmCl in the buffer was measured with a refractometer (Bausch and Lomb, Rochester, NY) using the method described by Nozaki [23].

### *Mass Spectrometry*

The MALDI matrix was prepared as a saturated SA solution in an aqueous buffer containing 0.1% TFA and 45% ACN. Internal standards (myoglobin and trypsin inhibitor) were added directly to the matrix solution. Mass spectra were acquired on one of two different MALDI mass spectrometers, including a Voyager Biospectrometry Workstation from PerSeptive Biosystems (Framingham, MA) and an Ultraflex II TOF/TOF from Bruker Daltonics (Billerica, MA). The Voyager instrument was employed in the initial screening experiment, and the Ultraflex instrument, which was not available for the project during the initial screening, was employed in the rescreening that was performed as part of the two-tier strategy (see below). Other than the Ultraflex's ability to collect data more rapidly, the performance characteristics (e.g., mass accuracy, precision, and resolution) were comparable.

All mass spectra were acquired in the linear and positive ion modes. Spectra collected with the Voyager instrument were a sum of 32 laser shots from a nitrogen laser operating at ~3 Hz and were collected using the following instrument parameters: an acceleration voltage of 25 kV, a grid voltage of 22.3 kV, a guide wire voltage of 37.5 V, and a delay time of 300 ns. Spectra collected with the Ultraflex instrument were a sum of 100 laser shots from a Nd:YAG laser operating at 100 Hz and were collected using the following instrument parameters: an ion source 1 voltage of 25 kV, an ion source 2 voltage of 23.35 kV, a lens voltage of 6 kV, and a delay time of 130 ns.

The mass-to-charge ratio of the CypA peak was determined using either a Microsoft Excel macro program or a MATLAB (The MathWorks, Inc., Natick, MA) script. Both programs require the input of spectral data in the form of a text file that contains *m/z* and intensity values for each data point in the spectrum. These

programs were developed in-house to perform the following three steps: (1) a 19-point floating average smoothing of the data, (2) a two-point calibration using the two internal standards, and (3) a center of mass calculation for the CypA peak. Each program uses the  $[M + H]^+$  ion signals of the two calibrants in each spectrum to generate a linear calibration equation that is then applied to the CypA ion signal. Ultimately, the center of mass of the CypA peak is calculated using a weighted average.

### *Library Screening*

Assay conditions (i.e., the H/D exchange time and denaturant concentration) for the initial library screening were chosen such that CypA ligands with  $K_d$  values in the low micromolar range would be selected. An exchange time of 35 min and a GdmCl concentration of 1.5 M were used here to select compounds with  $K_d$  values  $\leq 10 \mu\text{M}$ . In theory, the H/D exchange time and denaturant concentration can be tuned to select tighter (or weaker) binding ligands. However, the goal of this study was to select for ligands with  $K_d$  values  $\leq 10 \mu\text{M}$  (see the Results and Discussion section). In the assay, 1  $\mu\text{L}$  of a 1 mM solution of each library ligand (in DMSO) was combined with 9  $\mu\text{L}$  of the deuterated exchange buffer (which contained 1.5 M GdmCl). A 1- $\mu\text{L}$  aliquot of a protonated stock solution of CypA (~100  $\mu\text{M}$ ) was added to the resulting 10- $\mu\text{L}$  volume of ligand- and GdmCl-containing deuterated exchange buffer, and the H/D exchange reaction with CypA was allowed to proceed for 35 min. Each H/D exchange reaction was quenched by adding 1  $\mu\text{L}$  of the exchange reaction to 9  $\mu\text{L}$  of ice-cold matrix solution (see above). Finally, 1 to 2  $\mu\text{L}$  of the quenched reaction/matrix solution was spotted onto a MALDI target. Mass measurements from five replicate MALDI mass spectra were averaged, and this average was used to calculate the  $\Delta\text{mass}$  value. The standard deviations of the mass measurements were typically ~6 Da, or ~10% of the  $\Delta\text{mass}$  values. To limit the effects of differential back exchange (i.e., the loss of varying numbers of amide deuterons) from sample to sample, only 12 samples including 10 library compounds, one positive control (CsA,  $K_d \sim 30\text{--}200 \text{ nM}$  [13–17]), and one negative control (DMSO) were analyzed at a time. H/D exchange reactions were staggered at 20-min intervals to allow for nearly continuous data collection (i.e., MALDI-MS spectra were acquired for each sample set while the subsequent set was undergoing H/D exchange).

### *Data Analysis*

A 9-point central moving average was used to smooth the control data. This calculation involved an unweighted average of each data point with the four previous and four subsequent data points. This smoothing step helped to adjust for day-to-day variations in experimental conditions (i.e., temperature and humid-

ity) and helped to correct for differences in the degree of back exchange for each sample set. The averaged values were used to determine a unique cutoff value for each set of 10 library members.

The moving average strategy described above was used to calculate an average  $\Delta\text{mass}_{\text{av}}$  value for the negative control (i.e., a  $\Delta\text{mass}_{\text{av}}$  value) in each set of 10 library compounds, and this  $\Delta\text{mass}_{\text{av}}$  value was used to determine the cutoff for the corresponding dataset. The cutoff was determined by subtracting a specified number of standard deviations (i.e., 3.0, 2.5, or 2.0 standard deviations) from the  $\Delta\text{mass}_{\text{av}}$  value. A second MATLAB program, also developed in-house, was used to identify samples with  $\Delta\text{mass}$  values less than or equal to the cutoff value, and these samples were classified as hits.

$Z'$ -factors were calculated using eq 4 as described in reference [24].

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|} \quad (4)$$

In eq 4,  $\sigma_{c+}$  is the standard deviation of the positive control,  $\sigma_{c-}$  is the standard deviation of the negative control,  $\mu_{c+}$  is the mean of the positive control, and  $\mu_{c-}$  is the mean of the negative control.

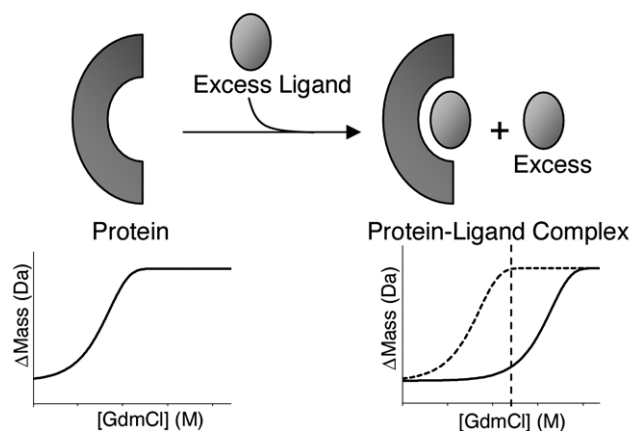
### Two-Tier Screening Strategy

Preliminary hits from the assay were subjected to an additional single-point SUPREX analysis as part of a two-tier screening strategy. The conditions of the re-screening experiment were similar to those of the initial screening, except that  $C_{18}$  ZipTips (Millipore, Billerica, MA) were used to concentrate and desalt the samples before single-point SUPREX analysis. The incorporation of ZipTips into the SUPREX protocol has been reported previously [25]. Briefly, this process involves quenching the H/D exchange reaction with 1  $\mu\text{L}$  of a 10% TFA solution and extracting the proteins from the H/D exchange buffers using ZipTips that were pre-equilibrated according to the manufacturer's instructions. Bound proteins were then washed with a solution of 0.1% TFA and eluted with the matrix solution (saturated SA in 50% ACN, 0.1% TFA, 49.9% water) directly onto the MALDI target. Internal standards were incorporated into the matrix solution before the elution step. This concentration and desalting step was not essential, but it generally improved the signal-to-noise ratio of the CypA peak.

## Results and Discussion

### General Strategy

The single-point SUPREX protocol used for the screening assay in this work relies on the ability of SUPREX to detect ligand binding (Figure 1). In a full SUPREX analysis, aliquots of a fully protonated protein solution are diluted into a series of deuterated exchange buffers



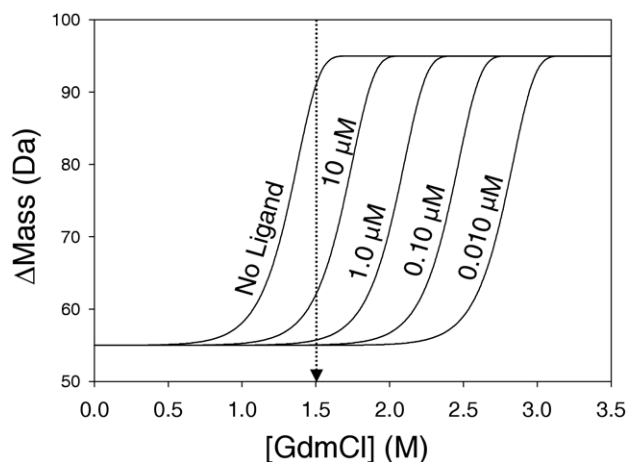
**Figure 1.** SUPREX-based detection of protein–ligand binding. A schematic representation of SUPREX curves expected in the absence and presence of ligand is shown in the bottom half of the panel. The dashed vertical line represents an appropriate denaturant concentration at which to perform a single-point SUPREX analysis. At this concentration, the separation between the  $\Delta\text{mass}$  values expected for the protein in the absence and in the presence of the ligand is largest. Shown in the top half of the panel is a schematic representation of the experimental conditions used to generate the theoretical SUPREX curves shown in the bottom half of the panel.

containing various concentrations of a chemical denaturant. The denaturant serves to shift the protein folding equilibrium toward the unfolded state. At higher concentrations of denaturant, a larger fraction of the protein population is unfolded, resulting in an increased uptake of deuterons and thus an increase in mass. This mass change (i.e.,  $\Delta\text{mass}$ ) is monitored by MALDI-MS. When the protein interacts with a ligand, the protein is stabilized, and a greater concentration of denaturant is required to unfold the protein. Thus, in the presence of a ligand, the midpoint of the SUPREX transition shifts toward higher denaturant concentrations.

SUPREX curves such as those shown in Figure 1 can be used to derive binding free energies and to determine solution phase dissociation constants ( $K_d$  values) with reasonably high accuracy and precision. However, it would be relatively time consuming and would require large amounts of protein to perform full SUPREX analyses for every member of a chemical library. The analysis time and amount of protein required can be significantly reduced by using a single-point SUPREX protocol in which a  $\Delta\text{mass}$  value is recorded at a single denaturant concentration for the protein in the presence of each library compound. If appropriate conditions (i.e., denaturant concentration and H/D exchange time) are chosen, the  $\Delta\text{mass}$  value will be low in the presence of a binding compound and high in the presence of a nonbinding compound (see dashed line in Figure 1).

The denaturant concentration and H/D exchange time used in this work were selected before the assay by constructing a series of theoretical SUPREX curves (see





**Figure 2.** Theoretical SUPREX curves for CypA in the absence and in the presence of four hypothetical ligands with various binding affinities. The arrow denotes the GdmCl concentration used in the single-point SUPREX analyses performed in this work.

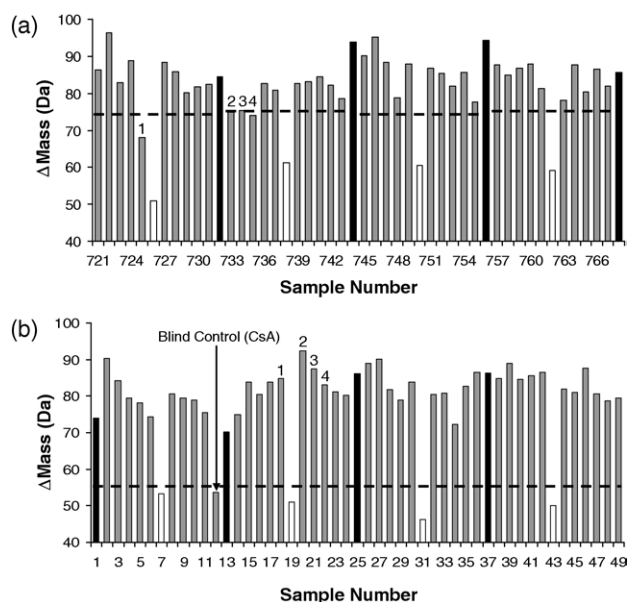
Figure 2) based on the known SUPREX behavior of CypA using GdmCl as the denaturant [17]. The effect of ligand binding on the SUPREX curve for CypA is illustrated in Figure 2, which shows a series of theoretical curves for CypA in the absence and presence of several hypothetical ligands with different binding affinities. A detailed description of the procedure used to generate the theoretical SUPREX curves is provided in the Experimental section. Using the theoretical curves in Figure 2, we chose a GdmCl concentration of 1.5 M and an exchange time of 35 min to allow for the selection of compounds in the library that bind CypA with a  $K_d \leq 10 \mu\text{M}$ . We reasoned that ligands with  $K_d$  values up to  $\sim 10 \mu\text{M}$  would be useful lead compounds in our search for CypA-targeting lung cancer therapeutics and imaging agents.

One advantage of using single-point SUPREX is that the assay is flexible and can be adjusted to select for a wide range of  $K_d$  values. Longer H/D exchanges times and/or higher denaturant concentrations could be used to select for tighter binding ligands. In theory, there is no upper or lower limit to the  $K_d$  value selection in single-point SUPREX. However, in practice the upper limit is defined by the protein and ligand concentrations that are experimentally accessible. For example, one would need to increase the protein and ligand concentrations tenfold compared with those used in this work to select ligands with  $K_d$  values  $>100 \mu\text{M}$ . The lower  $K_d$  value limit is not generally subject to experimental limitations. We also note that in a given single-point SUPREX selection, tighter binding ligands do not necessarily yield different  $\Delta\text{mass}$  values than weaker binding ligands, as long as the ligands have  $K_d$  values within the selection range. For example, in our selection of CypA ligands with  $K_d$  values  $\leq 10 \mu\text{M}$ , ligands with  $K_d$  values of 1.0, 0.10, and  $0.010 \mu\text{M}$  would all be expected to yield the same 55 Da  $\Delta\text{mass}$  value (see Figure 2).

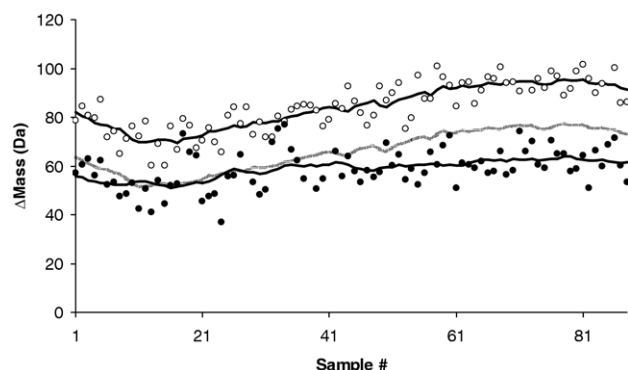
## Hit Identification

The 880 compounds in the Prestwick Chemical Library were individually screened for binding to CypA using single-point SUPREX. Representative data from the screen is shown in Figure 3a. Note that a positive control (CsA) and a negative control (DMSO) were analyzed with every set of 10 ligands in the library.  $\Delta\text{Mass}$  values were determined for all compounds in the library except for 18 compounds, all of which appeared to suppress the MALDI ion signal for CypA. Shown in Figure 4 are the  $\Delta\text{mass}$  values obtained for the positive and negative controls that were analyzed over the course of the screening, which took place in  $\sim 5$ -h time blocks spread over a period of several weeks. The control data were not constant over the time course of the experiment. Figure 4 shows a random scatter to the  $\Delta\text{mass}$  values obtained for the controls as well as a clear trend in the  $\Delta\text{mass}_{\text{av}}$  values that were calculated from the control data.

The random scatter of the points in Figure 4 can be explained by the precision of the molecular weight determinations of the CypA in this work. Standard deviations of replicate molecular weight determinations using our MALDI readout were consistently  $\sim 6$  Da, whether the protein was protonated, deuterated in the presence of ligand, or deuterated in the absence of



**Figure 3.** (a) Representative data from the initial screen performed using the single-point SUPREX assay. Positive and negative controls are shown as open and filled bars, respectively. The standard deviations for the mass measurements were typically  $\leq 10\%$  of the  $\Delta\text{mass}$  values. (b) The rescreening data collected on the preliminary 41 hits identified in the initial screen using a 3.0 standard deviation cutoff. The dotted lines in both panels represent the 3.0 standard deviation cutoff, which varied slightly for each dataset in panel (a) according to the moving average. The numbered bars represent compounds that were selected as hits in the initial screening but not in the rescreening, and the arrow indicates the only hit (i.e., the blind control) that was identified using the two-tier strategy.



**Figure 4.** Summary of the positive (filled circles) and negative (open circles) control data. The upper and lower solid black lines represent the moving average ( $\Delta\text{mass}_{\text{av}}$ ) of the negative and positive control data, respectively (see the Experimental section). The gray line depicts the moving cutoff value, which was set at three standard deviations below the moving average of the negative control data.

ligand. This is consistent with the mass accuracy expected for our MALDI-TOF instrument ( $\sim 200$  to  $400$  ppm). The changes observed in the  $\Delta\text{mass}_{\text{av}}$  values over the course of the experiment are most likely due to variations in the degree of back exchange for each sample set. This variability in the extent of back exchange most likely resulted from day-to-day fluctuations in the temperature and humidity levels of the laboratory and from variations in the timing of sample preparation/data collection.

The variability in the control data made the selection of a constant cutoff value impractical, and it prompted the development of a new strategy for processing single-point SUPREX data that involved the use of a moving average (i.e.,  $\Delta\text{mass}_{\text{av}}$ ) to calculate cutoff values. This moving average strategy was particularly useful in correcting for the day-to-day and set-to-set variations in back exchange that we observed. Initially, a cutoff value of 3 standard deviations below the  $\Delta\text{mass}_{\text{av}}$  value of the negative controls (Figure 4) was used to select hits. This resulted in the selection of 41 preliminary hits.

The 41 preliminary hits were rescreened in a second single-point SUPREX analysis (Figure 3b). Four positive and four negative controls were also included in this analysis. In the rescreening, the four negative control  $\Delta\text{mass}$  values were averaged to generate a  $\Delta\text{mass}_{\text{av}}$  value, and the cutoff was determined by multiplying the standard deviation of the negative control  $\Delta\text{mass}$  values by 3 and subtracting this product from the  $\Delta\text{mass}_{\text{av}}$  value of the negative control (79 Da). This calculation resulted in a cutoff of 55 Da, and application of this cutoff resulted in a single hit. The hit was the blind control, CsA. It is not surprising that no new hits were identified in a library of this size; a conservative estimate of the average rate of confirmed HTS hits is reported to be  $<0.1\%$  [26, 27]. Thus, a library of at least 1000 compounds is generally required to identify just one confirmed hit.

## Throughput

Assay data were collected during time periods of  $\sim 5$  h per day over a period of several weeks. In total, it took  $\sim 3000$  min, or about 3 min/ligand, to acquire the mass spectra for all library compounds and controls using the Voyager mass spectrometer that was exclusively employed in the initial screen. Extraction of the  $\Delta\text{mass}$  data from the  $>5000$  mass spectra recorded in this work took a total of  $\sim 300$  min, and subsequent analysis of the  $\Delta\text{mass}$  values to ascertain which compounds were hits was accomplished in  $\sim 30$  min. However, we estimate that the data analysis time could have been reduced approximately threefold had we exclusively used the MATLAB program for extracting  $\Delta\text{mass}$  values from the spectra. The MATLAB program was developed during the course of this work and was only employed in the hit validation component of the screening project. The Excel-based program, which is approximately tenfold slower than the MATLAB program, was employed in the early stages.

Several factors that impact the throughput of single-point SUPREX include the exchange time, the size of the sample set, and the rate at which MALDI mass spectra can be collected. Long exchange times require a larger investment of time before mass spectral analysis can begin. However, as long as H/D exchange reactions are staggered, the extra time is only invested once at the beginning of each assay. Increasing the size of each sample set could improve throughput as well, particularly if the application of a moving average is used to account for variations in back exchange. In this work, back exchange effects were minimized by keeping each sample set small (i.e., each set contained 10 library compounds, a positive control, and a negative control). This allowed for the collection of all spectra for each set within a short period of time (about 15 min). Small sample sizes also helped to decrease the time required to pipette samples onto the MALDI target, thereby minimizing the effects of differential back exchange. However, we note that  $\sim 50\%$  of the deuterons that exchanged into the protein during the H/D exchange reaction were consistently back-exchanged to protons during the MALDI sample preparation and analysis step, as evidenced by the  $\Delta\text{mass}$  values of the negative controls, which were typically  $\sim 80$  Da. This  $\Delta\text{mass}$  value represents  $\sim 50\%$  of the mass gain expected if all 167 amide protons in CypA were exchanged with solvent deuterons. The use of back exchange correction factors was investigated in this work (data not shown), but their use in the data analysis had no impact on the screening results.

We found that the rate-limiting step in our initial screen using the Voyager mass spectrometer was the time required to acquire the MALDI mass spectra. Due to the low repetition rate of the nitrogen laser ( $\sim 3$  Hz) and the slow sample positioning mechanisms ( $\sim 10$  s/sample) in the instrument,  $\sim 15$  min were required to collect the necessary mass spectra on each dataset of 12

samples. The use of a high throughput MALDI instrument, such as the Ultraflex instrument used in our rescreening experiments, could result in a dramatic increase in throughput. Such high throughput instruments have laser repetition rates of ~200 Hz and have fast sample positioning mechanisms (~2 s/sample). Additionally, the use of a high throughput MALDI instrument might allow for the analysis of larger datasets without significantly increasing the time that back exchange can occur within the mass spectrometer. We estimate that the use of a high throughput MALDI instrument for single point SUPREX analyses would increase the throughput of single-point SUPREX to <20 s/ligand. We note that the Ultraflex instrument was not used in the initial screening experiment in this work because it was not available for the project.

One additional approach for increasing the throughput of single-point SUPREX is to perform the analyses in a parallel fashion. In this study, each microtiter plate well contained only one library compound. However, single-point SUPREX is particularly well-suited for HTS applications because of its multiplex capabilities. Theoretically, multiple ligands could be placed in each well of the microtiter plate, allowing for the simultaneous screening of many library compounds at once. For example, the simultaneous screening of 10 library compounds per well has the potential to create a tenfold increase in throughput. One caveat to such multiplexing is that for every well that is identified as a hit, the compounds in the well must be individually retested to identify which compound(s) were responsible for the hit. Because of the need for additional analysis of each hit, the screen would need to be stringent enough to select for only the tightest binding ligands. The combined use of a high throughput MALDI and a multiplexing strategy is expected to produce up to a 100-fold increase in throughput compared to the 3 min/ligand observed here. This would make it possible to screen ~40,000 compounds in a 24-h day using single-point SUPREX.

### Efficiency

One goal of this work was to determine the efficiency of the single-point SUPREX protocol. The efficiency is defined here as the accuracy with which hits can be identified. One factor that influences efficiency is the choice of the cutoff value. The cutoff value used in this work was initially taken as 3 standard deviations below the  $\Delta\text{mass}_{\text{av}}$  values for each dataset. The use of three standard deviations is customary for high throughput screens [24], but we note that this value could be adjusted to make the cutoff more or less stringent. For example, the use of a cutoff value 2.5 standard deviations below the  $\Delta\text{mass}_{\text{av}}$  of the negative control resulted in the selection of 54 new preliminary hits in addition to the 41 preliminary hits selected using the three standard deviation cutoff. However, we note that a rescreening of the 54 new preliminary hits yielded no hits.

**Table 1.** Efficiency of single-point SUPREX

	False positives	False negatives
<b>Controls</b>		
3.0 SD cutoff	0/86 (0%)	15/85 (18%)
2.5 SD cutoff	0/86 (0%)	8/85 (9%)
<b>Library compounds (1-tier)</b>		
3.0 SD Cutoff	40/41 (98%)	—
2.5 SD Cutoff	94/95 (99%)	—
<b>Library compounds (2-tier)<sup>a</sup></b>		
3.0 and 3.0 SD cutoffs	0/1 (0%)	—
2.5 and 2.5 SD cutoffs	0/1 (0%)	—
2.5 and 3.0 SD cutoffs	0/1 (0%)	—

<sup>a</sup>In this category, the number of standard deviations used to calculate the cutoff values are listed for tiers 1 and 2, respectively.

Two important parameters used to define efficiency are the number of false positives and false negatives that appear in the screen. The control data provides one way in which to estimate the rate of false positives and false negatives. Summarized in Table 1 are the false positive and false negative rates observed if the 2.5 and 3.0 standard deviation cutoff values described above are applied to the positive and negative control data. It is noteworthy that the false positive rate of 0% for the controls is unchanged regardless of whether the cutoff is determined using 3.0 or 2.5 standard deviations. In contrast, the false negative rate is improved by a factor of two if the cutoff value is calculated using 2.5 instead of 3.0 standard deviations. Thus, the choice of cutoff value can be tuned to optimize false positive and false negative rates. The optimum cutoff values in this work were determined to be those calculated using 2.5 standard deviations in each tier (see below).

The false positive rate can also be estimated using the library members. Out of the 41 preliminary hits identified in the initial screen, only one was found to bind CypA when the 41 preliminary hits were rescreened. This gives a calculated false positive rate of 98%, which is high, especially considering that Hann and Oprea estimate a false positive rate of ~40% in typical pharmaceutical screens [26]. Interestingly, the false positive rate for the initial library selection is about the same (i.e., 99%) if the cutoff value is calculated using 2.5 standard deviations. The false negative rate cannot be easily calculated using the library members. However, our selection of CsA suggests that the false negative rate is low enough for the screen to be effective.

Based on the results obtained with the library members, the false positive rates observed in our initial screening of the 880-member library are quite high (see one-tier results in Table 1). One approach that can be used to reduce false positives in HTS assays is to use a two-tier selection strategy in which preliminary hits detected in an initial screen are subject to a rescreening [28]. Such a two-tier selection strategy is especially useful when the assay readout is subject to random error, as is the case for the MALDI-MS readout in the single-point SUPREX protocol. If such a two-tier selec-



tion strategy is used here to screen the 880-member library, the false positive rates observed for the single-point SUPREX protocol are reduced to 0%, regardless of how the cutoffs are calculated (i.e., using 3.0, 2.5, or a combination of 2.5 and 3.0 standard deviations).

Our results indicate that a two-tier selection strategy using 2.5 standard deviations to calculate the cutoff values in tier one was optimal in this work. The results for tier two were the same regardless of whether 2.5 or 3.0 standard deviations were used in the calculation of the cutoff value. As shown in Table 1, the choice of 2.5 standard deviations produced the optimal false positive and false negative rates for the controls, the lowest false positive rate in the library screening, and a reasonable number of preliminary hits (i.e., 10% of the 880 compounds in the library). We note that the use of 2.0 standard deviations to calculate the cutoff value was also investigated (data not shown); however, this resulted in a large number of preliminary hits (~25% of the 880 ligands in the library), an increased false positive rate based on the controls, and only a small improvement in the false negative rate (i.e., from 9% to 7%).

One drawback to the two-tier selection strategy is that extra time is required for the repeated analyses of false positives. Our results indicate that the time needed for such repeat analyses was relatively small (i.e., <10% of the time required for the preliminary screening) when the cutoff values in each tier were determined using 2.5 standard deviations. Moreover, the use of such a cutoff value only yielded a false negative rate of 9%. This suggests that the likelihood of *missing* a specific ligand in the library with a  $K_d$  value <10  $\mu$ M was low (i.e., <10%) and that the likelihood of *selecting* a specific ligand in the library with a  $K_d$  value <10  $\mu$ M was high (i.e., >90%).

Another parameter used to judge the efficiency of the assay is the  $Z'$ -factor, which can be calculated using eq 4 [24]. Using the  $\Delta\text{mass}_{\text{av}}$  and standard deviation for each control, it was possible to calculate  $Z'$ -factors for all of the control data points that were collected over the course of our screening. The  $Z'$ -factors ranged from -2.1 to 0.2, and the average  $Z'$ -factor was -0.7 over the entire screen. It is noteworthy that the control data improved over the course of the experiment. For example,  $Z'$ -factors ranged from -0.3 to 0.2 in the last 25% of the controls, and the average  $Z'$ -factor for these controls was 0.0. This improvement in the  $Z'$ -factor is attributed to an improvement in the timing of the H/D exchange reactions and mass spectrometry analyses, which decreased the variation in back exchange between data-sets. It has been noted that HTS assays with  $Z'$ -factors  $\geq 0$  are generally useful for screening large (e.g., 100,000-member) combinatorial libraries [24]. Our results suggest that with optimized timing of the H/D exchange reactions and mass spectrometry analyses the single-point SUPREX protocol is capable of screening such libraries for binding to CypA (i.e.,  $Z'$ -factors of 0 can be obtained).

The  $Z'$ -factors in this work were significantly lower than the  $Z'$ -factor calculated in the single-point SUPREX proof-of-concept study involving the S-protein ( $Z' = 0.77$ ) [10]. The discrepancy is most likely due to the difference in the masses of the target proteins. CypA has a mass of ~18.2 kDa, while the mass of the protein used in the proof-of-concept study (S-protein) was only ~11.5 kDa. Since the absolute precision of MALDI-TOF mass spectrometers decreases as protein mass increases, our mass measurements of CypA were less precise (i.e., standard deviations of ~6 Da) than our mass measurements of S-protein (i.e., standard deviations of ~2 Da). Higher standard deviations can result in lower  $Z'$ -factors. This makes the use of single-point SUPREX more challenging for larger proteins. However, the  $Z'$ -factor does not solely depend on the standard deviations of the controls; it also depends on the difference between the average positive and negative control values. Thus, the  $Z'$ -factor will also be influenced by the amplitude of the target protein's SUPREX curve. Such amplitudes are hard to predict for a given protein as they are largely defined by the number of globally protected amide protons (i.e., protons that are protected by the folded structure of the protein). Because of this dependence on the amplitude of the SUPREX curve, we expect the  $Z'$ -factor to be different for every target protein. Thus, larger proteins that have SUPREX curves with large amplitudes may still be amenable to single-point SUPREX assays. The amplitude of a SUPREX curve tends to be slightly larger when short exchange times are used [9]; therefore, for larger proteins, the  $Z'$ -factor could potentially be improved slightly by adjusting the exchange time.

Another potential strategy for improving the efficiency of single-point SUPREX for larger proteins involves the incorporation of a protease digestion step into the protocol. Recently, we reported on the use of a rapid (<2 min) protease digestion step to facilitate the SUPREX analyses of large multidomain proteins [29]. In this SUPREX-protease digestion protocol the H/D exchange properties of the individual domains of a protein are defined when the domains are in the intact protein. However, the domain-level folding properties are evaluated using peptides generated in the protease digestion step. The standard deviations associated with peptide molecular weight determinations by MALDI-TOF mass spectrometry are typically smaller than those associated with intact protein molecular weight determinations. These smaller standard deviations could result in a significant improvement in the  $Z'$ -factor of single-point SUPREX assays. Thus, the future incorporation of a protease digestion step into the single-point SUPREX protocol may also improve the efficiency with which compounds can be screened for binding to large proteins using SUPREX.

## Conclusions

We have demonstrated that a single-point SUPREX protocol can be used to screen the 880-member Prestwick Chemical Library for binding to CypA. This



application of single-point SUPREX provides the first measure of the throughput of this screening assay, which was determined to be 3 min/ligand using a conventional MALDI mass spectrometer equipped with a nitrogen laser operating at 3 Hz. As part of this work, we found that the throughput of the single-point SUPREX protocol was largely limited by the time needed to acquire the mass spectra in the MALDI readout. Thus, the use of conventional MALDI mass spectrometers that have relatively slow sample positioning mechanisms and nitrogen lasers with relatively slow repetition rates can create a bottleneck in HTS projects utilizing the single-point SUPREX protocol. The use of high throughput MALDI mass spectrometers equipped with fast sample positioning mechanisms and high repetition rate lasers could potentially increase the throughput of the protocol to less than 20 s/ligand. The data generated in this work also permitted the best evaluation of the efficiency of single-point SUPREX to date, and the results suggest that a two-tier selection strategy is important for minimizing the false positive and false negative rates of the single-point SUPREX assay.

## Acknowledgments

The authors thank Steven D. Hopper for his data analysis advice and for coding the MATLAB programs. The authors acknowledge support for this work by an NIH grant to M.C.F. (R21-CA-127064). E.D.H. was supported in part by an NIH training grant from the Duke University Center for Biomolecular and Tissue Engineering. P.L.R. was supported by a fellowship from the NIH (R01-GM-061680-S1).

## References

- Papac, D. I.; Shahrokh, Z. Mass Spectrometry Innovations in Drug Discovery and Development. *Pharm. Res.* **2001**, *18*, 131–145.
- Shin, Y. G.; van Breemen, R. B. Analysis and Screening of Combinatorial Libraries Using Mass Spectrometry. *Biopharm. Drug Dispos.* **2001**, *22*, 353–372.
- Sussmuth, R. D.; Jung, G. Impact of Mass Spectrometry on Combinatorial Chemistry. *J. Chromatogr. B* **1999**, *725*, 49–65.
- Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G. A. Quantitative, High-Throughput Screen for Protein Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8296–8301.
- Ma, L. Y.; Fitzgerald, M. C. A New H/D Exchange- and Mass Spectrometry-Based Method for Thermodynamic Analysis of Protein-DNA Interactions. *Chem. Biol.* **2003**, *10*, 1205–1213.
- Powell, K. D.; Fitzgerald, M. C. Accuracy and Precision of a New H/D Exchange- and Mass Spectrometry-Based Technique for Measuring the Thermodynamic Properties of Protein–Peptide Complexes. *Biochemistry* **2003**, *42*, 4962–4970.
- Powell, K. D.; Ghaemmaghami, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. A General Mass Spectrometry-Based Assay for the Quantitation of Protein–ligand Binding Interactions in Solution. *J. Am. Chem. Soc.* **2002**, *124*, 10256–10257.
- Powell, K. D.; Wang, M. Z.; Silinski, P.; Ma, L.; Wales, T. E.; Dai, S. Y.; Warner, A. H.; Yang, X.; Fitzgerald, M. C. The Accuracy and Precision of a New H/D Exchange- and Mass Spectrometry-Based Technique for Measuring the Thermodynamic Stability of Proteins. *Anal. Chim. Acta* **2003**, *496*, 225–232.
- Tang, L.; Hopper, E. D.; Tong, Y.; Sadowsky, J. D.; Peterson, K. J.; Gellman, S. H.; Fitzgerald, M. C. H/D Exchange- and Mass Spectrometry-Based Strategy for the Thermodynamic Analysis of Protein–ligand Binding. *Anal. Chem.* **2007**, *79*, 5869–5877.
- Powell, K. D.; Fitzgerald, M. C. High-Throughput Screening Assay for the Tunable Selection of Protein Ligands. *J. Combinat. Chem.* **2004**, *6*, 262–269.
- Campa, J.; Wang, M. Z.; Howard, B.; Fitzgerald, M. C.; Patz, E. F. Protein Expression Profiling Identifies Macrophage Migration Inhibitory Factor and Cyclophilin A as Potential Molecular Targets in Non-small Cell Lung Cancer. *Cancer Res.* **2003**, *63*, 1652–1656.
- Howard, B. A.; Furumai, R.; Campa, M. J.; Rabbani, Z. N.; Vujaskovic, Z.; Wang, X. F.; Patz, E. F. Stable RNA Interference-Mediated Suppression of Cyclophilin A Diminishes Non-Small-Cell Lung Tumor Growth in Vivo. *Cancer Res.* **2005**, *65*, 8853–8860.
- Handschumacher, R. E.; Harding, M. W.; Rice, J.; Drugge, R. J. Cyclophilin: A Specific Cytosolic Binding-Protein for Cyclosporin A. *Science* **1984**, *226*, 544–547.
- Harding, M. W.; Handschumacher, R. E. Cyclophilin, a Primary Molecular Target for Cyclosporine—Structural and Functional Implications. *Transplantation* **1988**, *46*, S29–S35.
- Holzman, T. F.; Egan, D. A.; Edalji, R.; Simmer, R. L.; Helfrich, R.; Taylor, A.; Burres, N. S. Preliminary Characterization of a Cloned Neutral Isoelectric form of the Human Peptidyl Prolyl Isomerase Cyclophilin. *J. Biol. Chem.* **1991**, *266*, 2474–2479.
- Liu, J.; Albers, M. W.; Chen, C. M.; Schreiber, S. L.; Walsh, C. T. Cloning, Expression, and Purification of Human Cyclophilin in *Escherichia coli* and Assessment of the Catalytic Role of Cysteines by Site-Directed Mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2304–2308.
- Wang, M. Z.; Shetty, J. T.; Howard, B. A.; Campa, M. J.; Patz, E. F.; Fitzgerald, M. C. Thermodynamic Analysis of Cyclosporin A Binding to Cyclophilin A in a Lung Tumor Tissue Lysate. *Anal. Chem.* **2004**, *76*, 4343–4348.
- Taylor, D. M. Radioiodinated Cyclosporin A: Preparation and Biodistribution. *Int. J. Appl. Radiat. Isot.* **1980**, *31*, 192–193.
- Bai, Y. W.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen-Exchange. *Prot. Struct. Funct. Genet.* **1993**, *17*, 75–86.
- Zhang, Y. Z. *Structural Biology and Molecular Biophysics*; Ph.D. Thesis, University of Pennsylvania, 1995.
- Schellman, J. A. Macromolecular Binding. *Biopolymers* **1975**, *14*, 999–1018.
- Glasoe, P. K.; Long, F. A. Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J. Phys. Chem.* **1960**, *64*, 188–190.
- Nozaki, Y. The Preparation of Guanidine Hydrochloride. *Methods Enzymol.* **1972**, *26*, 43–50.
- Zhang, J. H.; Chung, T. D. Y.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **1999**, *4*, 67–73.
- Powell, K. D.; Fitzgerald, M. C. Measurements of Protein Stability by H/D Exchange and Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Using Picomoles of Material. *Anal. Chem.* **2001**, *73*, 3300–3304.
- Hann, M. M.; Oprea, T. I. Pursuing the Leadlikeness Concept in Pharmaceutical Research. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255–263.
- Oprea, T. I. Current Trends in Lead Discovery: Are We Looking for the Appropriate Properties? *J. Comput. Aided Mol. Des.* **2002**, *16*, 325–334.
- Zhang, J.-H.; Wu, X.; Sills, M. A. Probing the Primary Screening Efficiency by Multiple Replicate Testing: A Quantitative Analysis of Hit Confirmation and False Screening Results of a Biochemical Assay. *J. Biomol. Screen.* **2005**, *10*, 695–704.
- Tang, L.; Roulhac, P. L.; Fitzgerald, M. C. H/D Exchange and Mass Spectrometry-Based Method for Biophysical Analysis of Multidomain Proteins at the Domain Level. *Anal. Chem.* **2007**, *79*, 8728–8739.